Supplementary Materials

"Experimental validation of the predicted binding site of Escherichia coli K1 outer membrane protein A to human brain microvascular endothelial cells; identification of critical mutations that prevent E coli meningitis."

Tod A. Pascal, Ravinder Abrol, Rahul Mittal, Ying Wang, Nemani V. Prasadarao and William A Goddard III

S.1. METHODS

S.1.1. Scanning the entire receptor for putative binding region

For a protein, there may be more than one binding region for one or more ligands. Thus our first step is to scan the entire protein for potential binding regions with no assumption on the binding site. The entire molecular surface of the predicted structure is mapped (1) and spheres representing the empty volume of the receptor are generated (currently using the Sphgen program in DOCK4.0 suite of programs (2). The entire set of receptor spheres is partitioned into ~30 to 50 overlapping regions and the ligand being docked is used to scan for the putative binding region. The consensus of ligand structures corresponding to the most energetically favorable sites is used to determine one or more putative binding region. For some ligands this scan of the entire receptor may yield two or three putative binding regions, with similar binding energies in each region. In such cases we merge the spheres of all the regions with similar binding regions and perform docking over this larger region.

S.1.2. Docking of chitobiose units to OmpA protein structures

The two chitobiose binding sites on OmpA (L and B) were determined above. The chitobiose unit was then docked into docked into each of these regions, with the top energy structure for each region retained. The binding is dominated by electrostatic energies (currently based on the RESP (3,4) scheme in this study, but on the Mulliken(5) scheme in our earlier study). The long range nature of the electrostatic interactions can lead to interaction energies that are highly sensitive to structural details remote from the binding regions. Thus, in our studies of ligands binding to G-Protein Coupled Receptors (GPCR), we showed that special approaches to neutralize the charged residues lead to dramatically better comparisons with experimental Kd values (6). The studies reported here use more traditional approaches but lead to very consistent interaction energies.

The GenMSCDock method was used to obtain ligand poses bound to OmpA. The method efficiently samples ligand conformational space and takes into account protein side-chain flexibility during the ligand binding process. The method is an extension of MPsim-Dock (7), and uses a Monte Carlo method to generate various possible ligand conformations in the field of the protein. The conformations are selected based on diversity of the conformations from each other to cover the conformational space. We call these conformations as "family heads" and they differ from each other by at least 0.6Å in CRMS (RMSD in coordinates). Next the energy of interaction of each family head with the protein is calculated, and about 10% of the good energy "family heads" are selected for further enrichment of these conformations. The enrichment is done by generating conformations using Monte Carlo procedure and selecting only those conformations that are close (within 0.6 Å CRMS) to the good energy family heads. A ligand conformation that is within 0.6 Å of the family head is known as a child of that family head. The enrichment cycle is performed until each chosen good energy family head has an average of at least 6 children. We then calculate the ligand protein interaction energy for all the children of each family head. The conformations (family heads and children) are sorted by energy and the best 50 ligand conformations are chosen, each of which is then taken through an optimum protein-side-chain placement using SCREAM side-chain optimization method (8). These structures are then conjugate-gradient minimized in two steps: first ligand-movable/ protein-fixed and second all-movable. Then the binding energy is calculated for these conformations.

This method was further improved by using a protein with bulky residues alanized to accommodate ligands. This replaces all bulky residues (Phe, Trp, Tyr, Val, Ile, and Leu) by Ala and all the docking is done to this bulky-alanized structure (other residues can be alanized if needed). The residues are dealanized to their original type, using SCREAM for each docking ligand conformation. It also has the option to neutralize charged residues that usually results in less noise in binding energies after correcting for the pKa of exposed residues. This allows for a better SAR (structure-activity relationship studies) as a smoother electrostatic potential in the binding region of the proteins leads to much smaller solvation energies and interaction energies both of which change much less for small changes in ligands.

S.2. RESULTS

In our earlier work, we reported computer docking studies of chitobiose epitopes interacting with OmpA (9). The authors note that these studies did not consider the lipid membrane bilayer and ignored the solvent (estimating solvation effects using an implicit model, based on continuum electrostatic theory, with no explicit water molecules). In the present work we find these techniques, while commonplace, are not adequate to capture the true nature of these systems and do not correlate well with our mutation experiments.

S.2.1. The stability of the mutated OmpA structures from computational studies

The CHRMM protein forcefield has been shown to provide a good description of secondary structure elements for a large set of experimental structures (10). An important question in such mutation experiments is whether our selected mutations would alter the OmpA fold (thereby destroying protein function). We tested this computationally, by performing 10ns of MD on the WT OmpA and on all ten mutants, without chitobiose bound. We characterized the effect of the mutations by analyzing the simulations to determine the changes in three characteristics:

1. the Cartesian Root Mean Square (CRMS) change in the structure
2. the Solvent Accessible Surface Area (SASA), a measure of the overall size exposed to solvent and
3. the Radius of gyration (R_{gyr}), a measure of the overall size of the protein (Table S2).

We find that none of the 10 sets of mutations led to a significant change in the structure, with the changes from the starting structure converging after 4ns. We found an average CRMS change of 1.6±0.4Å. The largest deviation was for the 1b mutants (2.19Å) and the smallest for the 2b mutant (1.16Å). Further, we find that the R_{gyr} of the mutants are similar to that of the WT (18.4±0.2Å), with R_{gyr} ranging from 18.4Å for 1b, 1c and 4c, to 18.9Å for 4a. In addition we find that the SASA ranges from 1% larger (4b) to 6% larger (2b) than the calculated WT value of 11196.9±103.1Å^2, (the NMR structure has SASA = 11043.1Å^2).

That the simulations find only small changes from WT gives us confidence in the accuracy of our methods and in the integrity of the mutated structures in mimicking the WT OmpA. We observed convergence to the equilibrated structure after 3.2 ns (Table S1). Most of the changes from the starting structure occurred in the loop region (CRMS = 3.30Å), with the CRMS of the B (barrel) binding site region < 0.5Å, while the CRMS of the L (loop) binding site regions are > 2.0Å. In particular loops 1, 2 and 3 (region L) fluctuate more than loop 4 (region B) during the MD.

S.3. CODE AVAILABILITY

The 2PT code is compatible with the LAMMPS (11,12) simulation engine. Interested parties can contact the authors (wag@wag.caltech.edu and tpascal@wag.caltech.edu) for access.

S.4. REFERENCES

1. Connolly, M. L. (1983) *Journal of Applied Crystallography* **16**, 548-558
2. Ewing, T. J. A., and Kuntz, I. D. (1997) *J. Comput. Chem.* **18**, 1175-1189
3. Bayly, C. I., Cieplak, P., Cornell, W., and Kollman, P. A. (2002) *The Journal of Physical Chemistry* **97**, 10269-10280
4. U. Chandra Singh, and Peter A. Kollman. (1984) *J. Comput. Chem.* **5**, 129-145
5. Mulliken, R. S. (1955) *J. Chem. Phys.* **23**, 1833-1840
6. Bray, J. K., and Goddard, W. A. (2008) *Journal of Molecular Graphics & Modelling* **27**, 66-81
7. Cho, A. E., Wendel, J. A., Vaidehi, N., Kekenes-Huskey, P. M., Floriano, W. B., Maiti, P. K., and Goddard, W. A. (2005) *J. Comput. Chem.* **26**, 48-71
8. Kam, V. W. T., and Goddard, W. A. (2008) *Journal of Chemical Theory and Computation* **4**, 2160-2169
9. Datta, D., Vaidehi, N., Floriano, W. B., Kim, K. S., Prasadarao, N. V., and Goddard, W. A. (2003) *Proteins* **50**, 213-221
10. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) *J. Phys. Chem. B* **102**, 3586-3616
11. Plimpton, S. J., Pollock, R., and Stevens, M. (1997) Particle-Mesh Ewald and rRESPA for Parallel Molecular Dynamics Simulations. in *Proc of the Eighth SIAM Conference on Parallel Processing for Scientific Computing*, Minneapolis, MN
12. Plimpton, S. (1995) *J. Comput. Phys.* **117**, 1-19

S.5. TABLES

Table S1: Properties of OmpA and its 10 mutants during dynamics (without chitobiose units bounds). Statistics are obtained over the last 10ns NPT MD trajectory. We present two measures of the heavy atom coordinate root means square deviations [CRMSD] (a) using the starting structure as the reference, to indicate how much the system relaxes from the initial minimized structure during dynamics and (b) using the equilibrated structure as a reference, to indicate convergence. Equilibration in all the structure is seen after 4ns of MD.

	SASA (Å^2)	**Rgyr (Å)**	**CRMSD (Å)**	
			Starting	***Equilibrated***
WT	11196.9±103.1	18.38±0.07	3.30±0.90	1.68±0.20
1a	11531.3±220.7	18.56±0.06	3.58±0.90	1.77±0.14
1b	11510.8±303.1	18.39±0.09	3.61±1.11	2.19±0.34
1c	11688.0±156.3	18.38±0.08	3.59±0.73	1.44±0.11
2a	11545.8±101.0	18.54±0.07	3.31±0.70	1.31±0.14
2b	11880.8±64.6	18.50±0.02	3.07±0.58	1.16±0.03
2c	11723.1±43.4	18.45±0.12	3.52±0.61	2.14±0.29
3a	11676.6±134.4	18.70±0.14	3.27±0.86	1.71±0.17
4a	11674.7±92.3	18.90±0.13	4.57±1.24	2.13±0.53
4b	11302.7±169.6	18.44±0.70	3.41±0.60	1.18±0.13
4c	11782.7±224.822	18.38±0.08	3.17±0.69	1.26±0.13

Table S1: Primers for site-directed mutagenesis.

Name	Sequence (5'-3')	Mutation
Omp-F1a	GGTTTCATCAACAAC ***GCTGCCGCG*** ACCCATGAAAACC	Loop 1a
Omp-R1a	GGTTTTCATGGGT ***CGCGGCAGC*** GTTGTTGATGAAACC	
Omp-F1b	ACAACAATGGCCCG ***GCCGCTGCA*** AACCAACTGGGCG	Loop 1b
Omp-R1b	CGCCCAGTTGGTT ***TGCAGCGGC*** CGGGCCATTGTTGT	
Omp-F1c	GGCTGGTCCCAG ***GCCGCTGCC*** ACTGGTTTCATCA	Loop 1c
Omp-R1c	TGATGAAACCAGT ***GGCAGCGGC*** CTGGGACCAGCC	
Omp-F2a	CGTATGCCGTACAAA ***GCCGCCGCTG****CA* AACGGTGCATAC	Loop 2a
Omp-R2a	GTATGCACCGTT ***TGCAGCGGCGGC*** TTTGTACGGCATACG	
Omp-F2b	CTGGTTAGGTCGT ***GCGGCGGCCGCA*** GGCAGCGTTGAAA	Loop 2b
Omp-R2b	TTTCAACGCTGCC ***TGCGGCCGCCGC*** ACGACCTAACCAG	
Omp-F3a	CAGACACTAAATCC ***GCCGCTGCT G***GTAAAAACCACGA	Loop 3a
Omp-R3a	CGTGGTTTTTACC *AG****CAGCGGC*** GGATTTAGTGTCTGC	
Omp-F4a	CAGTGGACCAACAAC ***GCCGCTGCC*** GCACACACCATCG	Loop 4a
Omp-R4a	CGATGGTGTGTGC ***GGCAGCGGC*** GTTGTTGGTCCACTG	
Omp-F4b	ACATCGGTGACGCA ***GCTGCCGCCGCC*** ACTCGTCCGGAC	Loop 4b
Omp-R4b	GTCCGGACGAGT ***GGCGGCGGCAGC*** TGCGTCACCGATGT	
Omp-F4c	CATCGGCACTCGT ***GCGGCTGC****C* GGCATGCTGAGC	Loop 4c
Omp-R4c	GCTCAGCATGCC ***GGCAGCCGC*** ACGAGTGCCGATG	

* Changed nucleotides are in bold and italics.

Table S3: Potential energy (ΔE^{md}), internal energy including zero-point energy corrections (ΔU^0), total entropy ($T\Delta S^0$) and Helmholtz free energy ($\Delta A^0 = \Delta U^0 - T\Delta S^0$) WT OmpA and 10 mutants calculated with 2PT method from MD trajectories. All energies in kcal/mol. Energies are referenced to the solvated chitobiose and OmpA structures.

	ΔE^{md}		ΔU^0		$T\Delta S^0$		ΔA^0	
	avg	±	avg	±	avg	±	avg	±
1a	-92.51	4.12	-5.49	4.96	-7.28	2.15	1.79	8.05
1b	-27.40	5.35	-2.60	5.30	-2.25	1.59	-0.35	6.46
1c	-13.64	4.14	-78.64	3.90	-33.69	1.86	-44.95	5.10
2a	-87.76	6.73	-227.76	5.26	-34.92	5.00	-192.84	6.58
2b	-148.15	9.46	-198.15	2.85	-69.91	1.37	-128.24	4.51
2c	-46.41	3.57	-86.41	7.30	-61.42	1.26	-24.99	7.48
3a	-261.65	6.63	-271.65	5.23	-84.73	5.26	-186.92	6.99
4a	-207.34	7.48	-67.34	5.19	-38.24	3.52	-29.10	5.97
4b	-26.33	3.38	-166.33	7.10	-99.38	1.36	-66.95	7.63
4c	-44.04	5.92	-144.04	5.14	-50.69	0.52	-93.35	6.61
WT	-334.14	3.82	-284.14	3.07	-78.12	7.92	-206.02	8.57

Table S4: Comparison of predicted and experimentally measured meningitis rates for OmpA and the 10 mutants considered in this study. Binding energies are relative to the solvated chitobiose and protein unit. There is a 90% correlation between the experimentally observed invasions and the relative binding free energy (ΔA^0). Considering only the internal energy with zero-point energy corrections (ΔU^0) leads to a 77% correlation. There is a poor 31% correlation with the MD energy (ΔE^{md}).

	Exp		**ΔE^{md}**		**ΔU^0**		**ΔA^0**	
	Avg.	±	Avg.	±	Avg.	±	Avg.	±
1a	1.54	0.00	35.43	4.45	1.76	90.35	-0.79	4.50
1b	1.54	1.54	10.49	19.51	0.83	203.88	0.15	1.80
1c	43.08	13.85	5.22	30.35	25.22	4.96	19.89	1.10
2a	81.54	15.38	33.61	7.67	73.04	2.31	85.31	3.41
2b	40.00	16.92	56.75	6.39	63.55	1.44	56.73	3.52
3a	75.38	7.69	100.22	2.53	87.12	1.92	82.69	3.74
4a	15.38	10.77	79.42	3.61	21.60	7.71	12.88	2.00
4b	20.00	10.77	10.09	12.84	53.34	4.27	29.62	1.14
4c	35.38	9.23	16.87	13.43	46.19	3.57	41.29	7.08
WT	100.00	7.69	127.98	1.14	91.13	1.08	91.14	4.16
R^2			0.31		0.77		0.90	
2c*			17.78	1.69	27.71	2.44	11.06	3.0

*the 2c mutant was not determined experimentally, we present a prediction here that can be tested

S.6. FIGURES

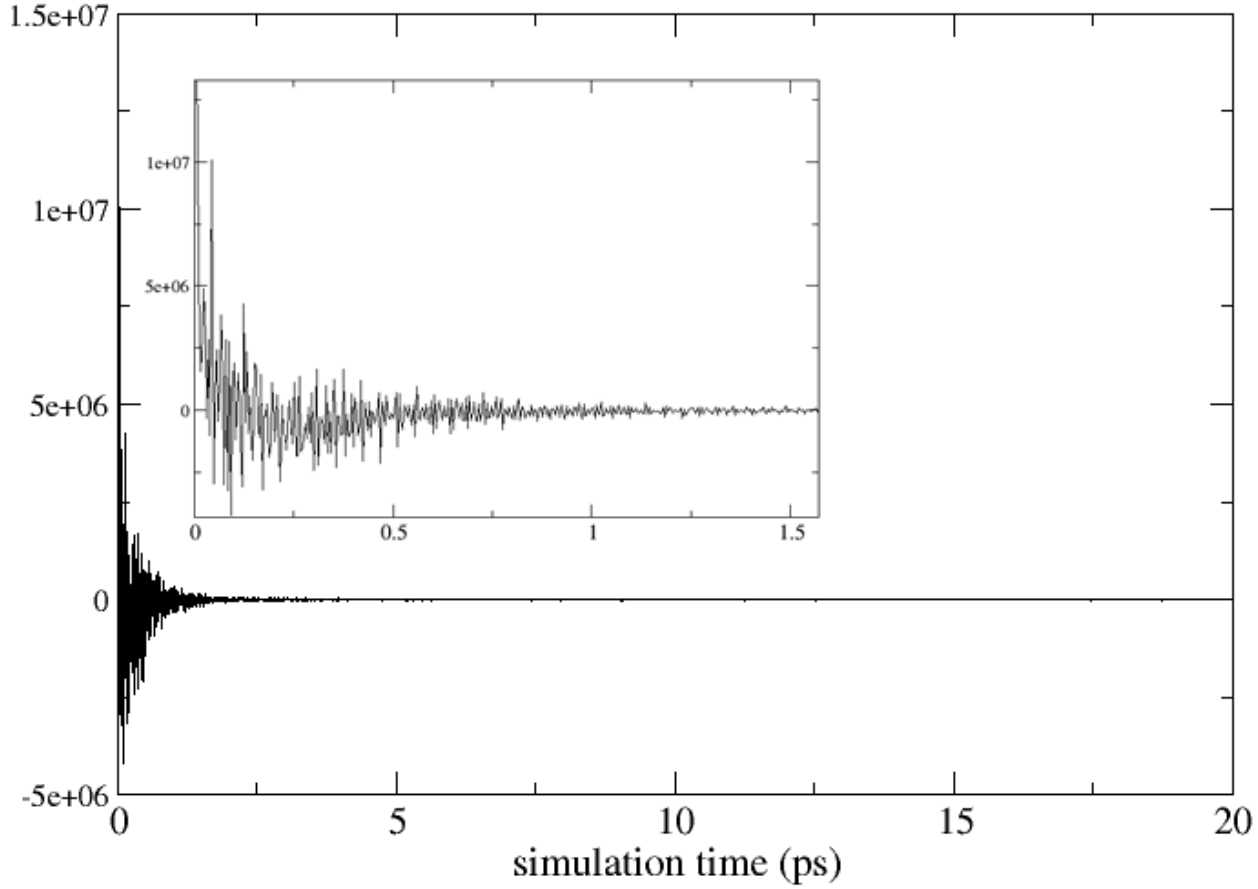

Figure S1: Velocity autocorrelation function (VAC) for WT OmpA from 2ns NPT MD, following the 30ns annealing dynamics. The density of states (DoS) is obtained from the Fourier Transform of this VAC. The VAC decays to zero after 5ps. Inset: magnification of the first 1.5 ps.

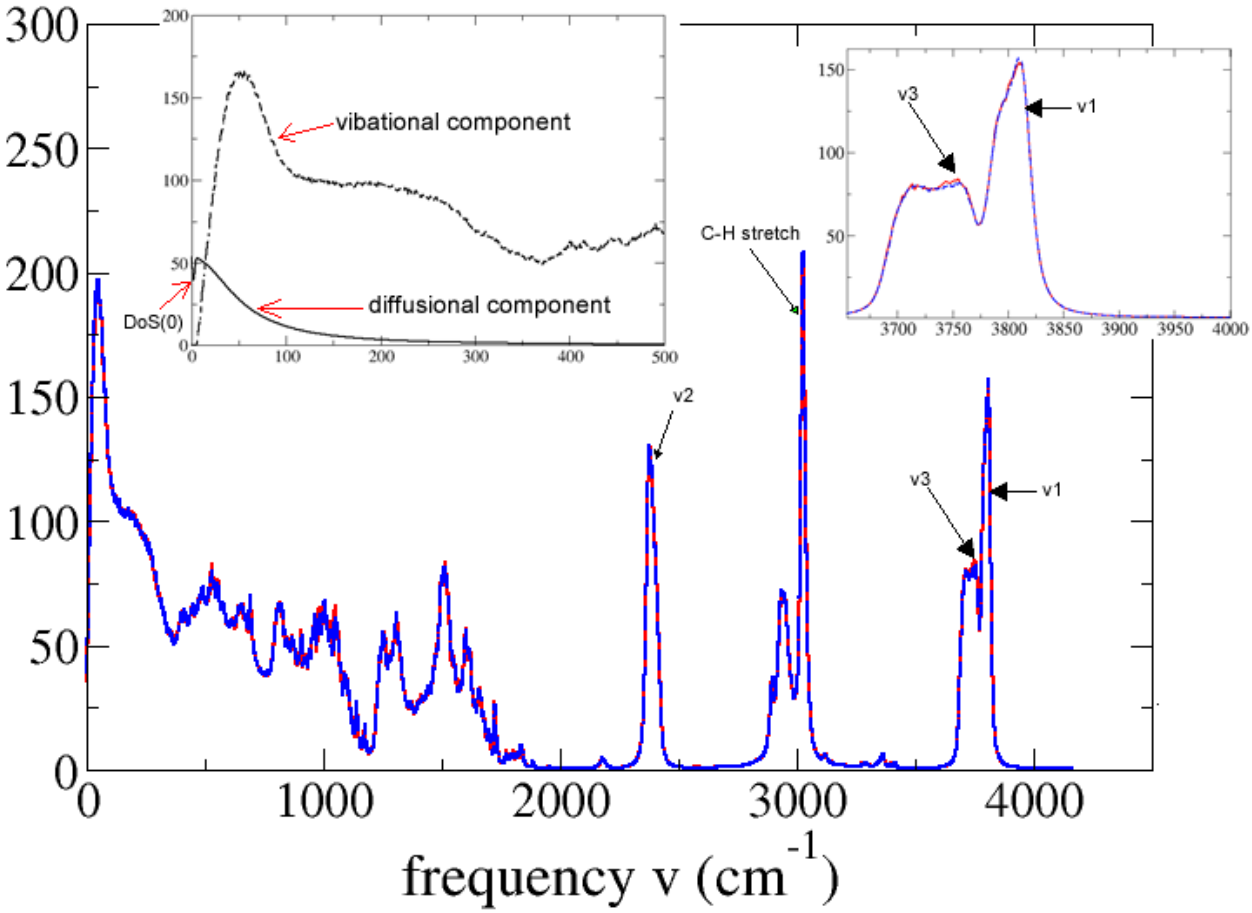

Figure S2: DoS of WT (solid red) and 1b mutant (dashed blue) calculated at 6ns during NPT MD. The characteristic vibrational frequencies are labeled: O-H stretch (water symmetric v1: 3806 cm-1, water asymmetric v_3: 3750 cm^{-1}), H-O-H bend v_2 (2375 cm^{-1}) and C-H stretch (3025 cm^{-1}). The v_2 frequencies are the only ones not close to experiment (1600 cm^{-1}) and is a deficiency of the water forcefield used (the force constant for the TIP4P H-O-H bend is 200 kcal/mol, twice the value needed to reproduce the experimental frequencies). Insets: 1) upper left - low frequency modes of WT, showing the decomposition of the total DoS into the diffusional and vibrational (solid) components. *DoS(0)* (the value of the DoS at *v* = 0) is 53.46 cm/mol, which is related to the self diffusion constant D_0=0.0973, according to equation (6). For the WT, we obtain f=0.1844 which corresponds to 29610 of the 160839 total 3N modes. The 1a mutant has more entropy than the wild type, due to the more mobile chitobiose unit (these low frequency diffusional modes contribute most to the entropy). 2) upper right - O-H stretching modes. The broadening and splitting of the O-H band is representative is caused by the hydrogen bonding in the system.

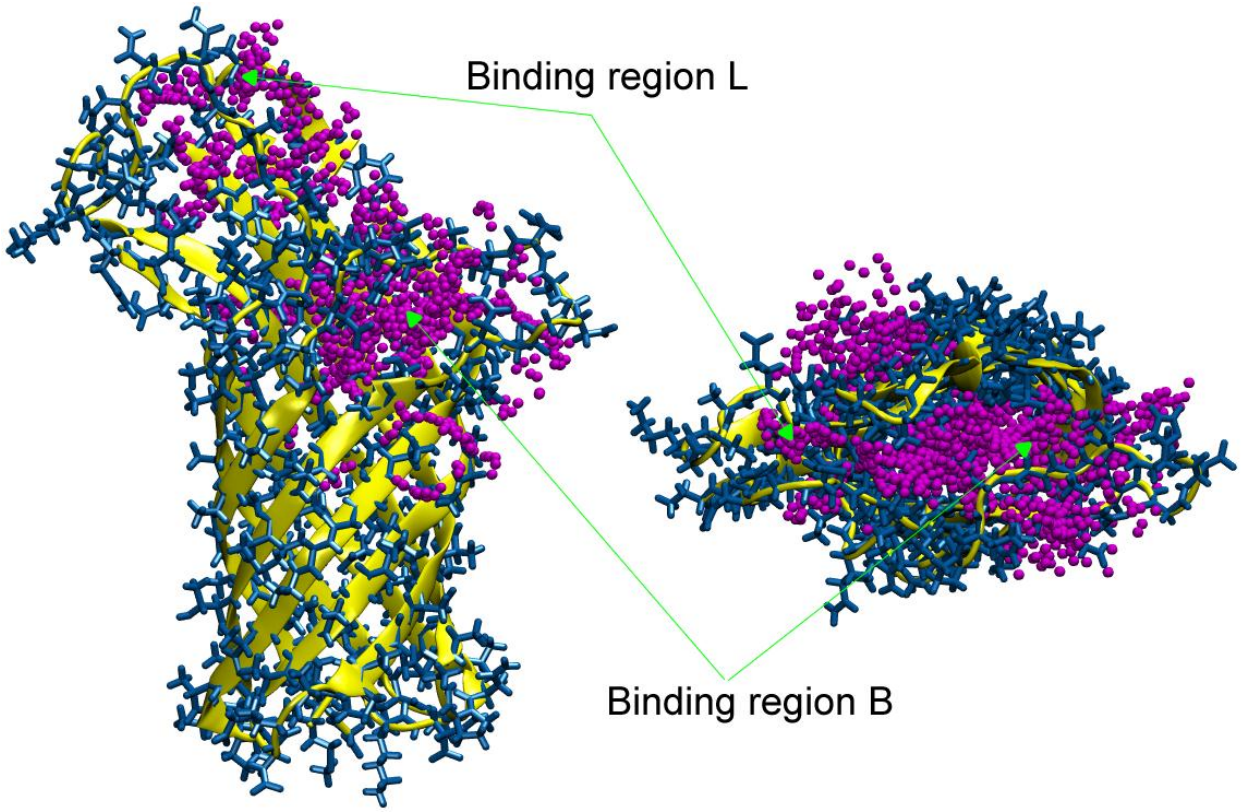

Figure S3**:** Volume filling spheres, with the chitobiose binding site (L and B) shown.

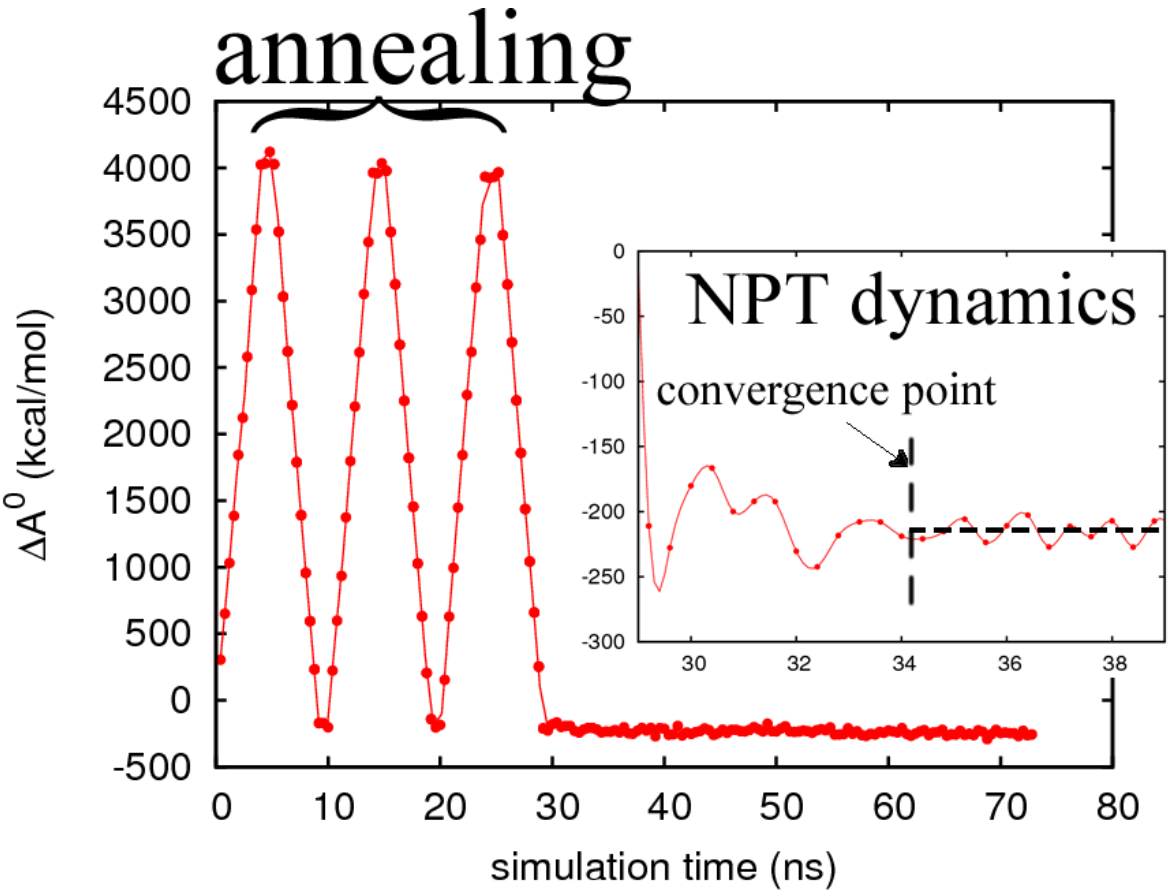

Figure S4: Relative binding free energy profile of WT OmpA with bound chitobiose during dynamics. The free energy is taken relative to separated chitobiose and WT OmpA structures. The binding free energy is -206 ± 8.6 kcal/mol and convergence is observed after 4ns of NPT MD, following the 30ns of annealing dynamics (74ns of total simulation time).

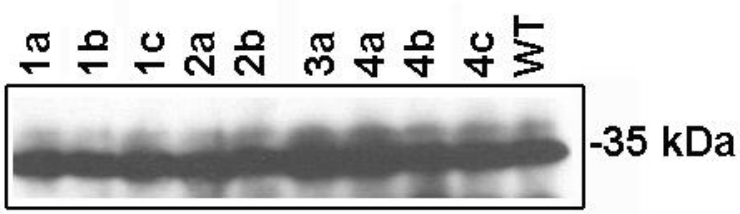

Figure S5: Western blot analysis of OmpA proteins from *E. coli* mutants. *E. coli* strains were grown overnight in Luria broth, washed and outer membrane proteins isolated. Equal amounts (10 μg per lane) was loaded and separated on SDS-polyacrylamide gel. The proteins were then transferred to a nitrocellulose membrane and immunoblotted with anti-OmpA antibodies. The bound antibody was probed with HRP-coupled secondary antibody followed by ECL reagent.